

Solution 1: Kullback-Leibler Divergence and model misspecification

(a) The Kullback-Leibler Divergence is defined as:

$$\begin{aligned} D(g, f_\theta) &= \int_{-\infty}^{\infty} \log\left(\frac{g(x)}{f_\theta(x)}\right) g(x) dx \\ &= \underbrace{\int_{-\infty}^{\infty} \log(g(x)) g(x) dx}_{(a)} - \underbrace{\int_{-\infty}^{\infty} \log(f_\theta(x)) g(x) dx}_{(b)} \end{aligned} \quad (1)$$

As we are looking for the set of parameters θ that minimizes $D(g, f_\theta)$, we know the following:

- (a) does not depend on θ , and can be considered as a constant.
- To minimize $D(g, f_\theta)$ is equivalent to maximize (b)

Using the definition of the normal distribution:

$$\begin{aligned} (b) &= \int_{-\infty}^{\infty} \log(f_\theta(x)) g(x) dx \\ &= \int_{-\infty}^{\infty} \left(\log\left(\frac{1}{\sqrt{\sigma^2 2\pi}}\right) - \frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right) g(x) dx \\ &= \log\left(\frac{1}{\sqrt{\sigma^2 2\pi}}\right) \underbrace{\int_{-\infty}^{\infty} g(x) dx}_1 - \int_{-\infty}^{\infty} \frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} g(x) dx \\ &= -\log\sqrt{\sigma^2 2\pi} - \underbrace{\int_{-\infty}^{\infty} \frac{1}{2} \frac{x^2 - 2x\mu + \mu^2}{\sigma^2} g(x) dx}_{(c)} \end{aligned} \quad (2)$$

Solving the component (c) in the equation 2 we get:

$$\begin{aligned} (c) &= -\frac{1}{2\sigma^2} \underbrace{\int_{-\infty}^{\infty} x^2 g(x) dx}_{\mathbb{E}_g(x^2) = \text{Var}_g(x) + \mathbb{E}_g[x]^2} + \frac{2\mu}{2\sigma^2} \underbrace{\int_{-\infty}^{\infty} x g(x) dx}_{\mathbb{E}_g[x]} - \frac{\mu^2}{2\sigma^2} \underbrace{\int_{-\infty}^{\infty} g(x) dx}_{=1} \\ &= -\frac{\sigma_0^2 + \mu_0^2}{2\sigma^2} + \frac{\mu\mu_0}{\sigma^2} - \frac{\mu^2}{2\sigma^2} \end{aligned} \quad (3)$$

Now, using the results obtained in 2 and 3, we get the expression that we want to maximize:

$$(b) = -\log\sqrt{\sigma^2 2\pi} - \frac{\sigma_0^2 + \mu_0^2}{2\sigma^2} + \frac{\mu\mu_0}{\sigma^2} - \frac{\mu^2}{2\sigma^2} \quad (4)$$

To maximize 4, we derive the expression with respect to each parameter. We also need to do a second derivative to be sure that the point is a maximum.

First, we derive with respect to the mean parameter μ :

$$\frac{\partial(b)}{\partial\mu} = 0 - 0 + \frac{\mu_0}{\sigma^2} - \frac{\mu}{\sigma^2} \stackrel{!}{=} 0 \longrightarrow \mu_{opt} = \mu_0 \quad (5)$$

This value of μ is a possible maximum, we check the second derivative:

$$\frac{\partial^2(b)}{\partial^2\mu} = -\frac{1}{\sigma^2} < 0 \quad (6)$$

As the second derivative is less than 0 at any point, μ_{opt} minimizes the Kullback-Leibler divergence. We now derive with respect to the variance parameter σ^2 :

$$\begin{aligned} \frac{\partial(b)}{\partial\sigma^2} &= -\frac{1}{2\sigma^2} + \frac{\sigma_0^2 + \mu_0^2}{2\sigma^4} - \frac{\mu\mu_0}{\sigma^4} + \frac{\mu^2}{2\sigma^4} \\ &= -\frac{1}{2\sigma^2} + \frac{\sigma_0^2 + \mu_0^2 - 2\mu\mu_0 + \mu^2}{2\sigma^4} \\ &= -\frac{1}{2\sigma^2} + \frac{\sigma_0^2 + (\mu - \mu_0)^2}{2\sigma^4} \stackrel{!}{=} 0 \longrightarrow \sigma_{opt}^2 = \sigma_0^2 + (\mu - \mu_0)^2 \end{aligned} \quad (7)$$

This value of σ^2 is a possible maximum, we check the second derivative:

$$\begin{aligned} \frac{\partial^2(b)}{\partial^2\sigma^2} &= \frac{1}{2\sigma^4} - \frac{(\sigma_0^2 + (\mu - \mu_0)^2)}{\sigma^6} \\ \frac{\partial^2(b)}{\partial^2\sigma^2} \Big|_{\sigma^2=\sigma_{opt}^2} &= \frac{1}{2(\sigma_0^2 + (\mu - \mu_0)^2)^2} - \frac{1}{(\sigma_0^2 + (\mu - \mu_0)^2)^2} < 0 \end{aligned} \quad (8)$$

As the second derivative is less than 0 at the point we are looking, σ_{opt}^2 minimizes the Kullback-Leibler Divergence.